SWM : A CLASS OF CONVEX CONTRASTS FOR SOURCE SEPARATION

Frédéric Vrins, Michel Verleysen

Université catholique de Louvain
UCL Machine Learning Group
3, Place du Levant
1348 Louvain-la-Neuve, Belgium
{vrins,verleysen}@dice.ucl.ac.be

Christian Jutten

LIS (CNRS UMR 5083, INPG, UJF)
46, Avenue Félix Viallet
38031 Grenoble cedex, France
christian.jutten@inpg.fr

ABSTRACT

We derive a class of contrasts for blind source separation (BSS) to separate bounded sources (or more generally, finite sources), based on support width measures (SWM) of the marginal output distributions. These contrasts are shown to have no spurious local maxima, i.e. all the local maxima are relevant from the source separation point of view; they all correspond to non-mixing BSS solutions so that a gradient-ascent method can be used.

1. INTRODUCTION

Linear blind source separation (BSS) consists in recovering unobservable source signals $\mathbf{s} = [s_1, \ldots, s_n]$ only knowing linear mixtures of them $\mathbf{x} = [x_1, \ldots, x_n]$, or their whitened version $\mathbf{z} = \mathbf{V}\mathbf{x}$ ($\mathbf{V}$ is the whitening matrix). The aim of BSS is to linearly combine the $\mathbf{z}$ such that the output signals $\mathbf{y} \doteq \mathbf{U}\mathbf{z}$ are proportional to the original source signals $\mathbf{s}$, up to a permutation. The mixture and separation schemes become:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

$$\mathbf{y} = \mathbf{U}\underbrace{\mathbf{V}\mathbf{x}}_{\mathbf{z}} = \underbrace{\mathbf{U}\mathbf{V}\mathbf{A}}_{\mathbf{W}}\mathbf{s} \tag{2}$$

Under the $E\{\mathbf{s}\} = \mathbf{0}$ and $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{I}$ assumptions, $\mathbf{VA}$ is orthogonal. In dimension two ($n = 2$), an orthogonal matrix is fully determined by a single angle, called here the *mixing angle* ϕ. The product $\mathbf{VA}$ is then a pure rotation ($\det(\mathbf{VA}) = 1$) or roto-inversion (rotation with inversion: $\det(\mathbf{VA}) = -1$) matrix. One can freely assume that $\mathbf{VA}$ is a pure rotation matrix; the only consequence of this assumption is that the sign and permutation indeterminacies are then linked to each other. There is no loss of information to suppose $\det(\mathbf{VA}) = 1$. Hence, the mixing and prewhitening steps can be expressed as follows:

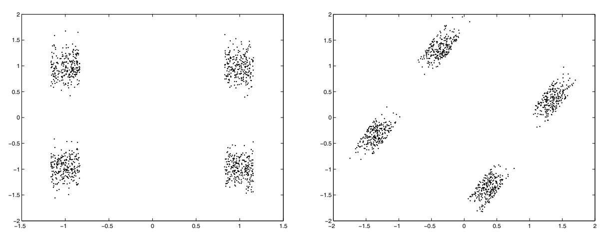
$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} . \tag{3}$$

Michel Verleysen is a Senior Research Associate of the Belgian National Fund for the Scientific Research (FNRS).

Furthermore, under the additional $E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{I}$ constraint, $\mathbf{U}$ also reduces to an orthogonal (assumed pure rotation) matrix, parametrized by a single *unmixing angle* φ. Hence, for $n = 2$, eq. 2 becomes:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos(\phi+\varphi) & \sin(\phi+\varphi) \\ -\sin(\phi+\varphi) & \cos(\phi+\varphi) \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \tag{4}$$

Hence, the BSS problem reduces to finding the unknown initial angle ϕ only knowing $\mathbf{z}$ and $\mathbf{y}$ by adjusting φ. Examples of scatter plots of $\mathbf{s}$ and $\mathbf{z}$ are illustrated in Fig. 1.



(a) s_1 vs s_2 (b) z_1 vs z_2

Fig. 1. Scatter plots (a) of the bimodal sources $\mathbf{s}$ and (b) of the whitened signals $\mathbf{z}$; $\phi = \pi/6$.

Let us define $\theta \doteq \phi + \varphi$. The angle ϕ is fixed, since $\mathbf{A}$ and $\mathbf{V}$ are constant matrices, but θ is unknown, and may vary through φ. Consequently, $\mathbf{W}(\theta)$ is non-mixing if and only if we have found *blindly* (ϕ is unknown) $\varphi = \varphi^\star$ such that $\varphi^\star = k\pi/2 - \phi$, $k \in \mathbb{Z}$.

If the s_i are mutually independent and if at most one of them has a Normal distribution, the well-known method of independent component analysis (ICA) is able to estimate the mixing angle ϕ. Several contrast functions $\mathcal{C}(\theta)$ have been derived to estimate the independence between variables [1]; their maxima correspond to satisfactory solutions of the BSS problem. In order to avoid an exhaustive search, most of non-algebraic ICA algorithms use a gradient ascent on $\mathcal{C}$ with respect to the angle φ (see [2] for a survey on algebraic methods). In order to avoid possible spurious maxima, all the local maxima of $\mathcal{C}$ must then appear for $\theta = k\pi/2$. A sufficient condition is to constraint the contrast to be convex for θ in each quadrant $\mathcal{Q}_p$.

In [3], Pham showed that the opposite of the *sum* of the log-marginal support widths of the output distributions is a global contrast for source separation. In this paper, we will derive a convex contrast, based on the support width of a *single* output distribution to separate bounded sources. Complementarily to Pham's approach, this contrast has theoretically no mixing maxima, and allows one to derive a deflation algorithm.

In the next section, the support width $||\Omega_{y_i}||$ of the output distribution p_{y_i} is related to the support width of the sources $||\Omega_{s_i}||$. Next, it is shown that $-||\Omega_{y_i}||$ is a convex contrast: each local maximum is an acceptable solution to the BSS problem, and the global maximum corresponds to the extraction of the source with minimum support width. A simple example illustrates this property in Section 3. It shows that using a well-defined support width measure, this contrast can be extended to so-called *finite source* signals.

2. SUPPORT WIDTH OF MIXTURE DENSITIES

In this section, the support Ω_y of $y = \alpha u + \beta v$ is related to the mixed variable supports $\{\Omega_u, \Omega_v\}$ and to the mixture coefficients α, β.

2.1. Bounded sources

Definition 1 (Support width of a bounded variable) *Let u be a one-dimensional random variable, bounded on $[l_u,r_u]$ where $l_u = \min(u)$ and $r_u = \max(u)$. The support of u is defined as $\Omega_u \doteq [l_u, r_u]$, where $-\infty < l_u < r_u < +\infty$. The support width of u is defined as $||\Omega_u|| \doteq |r_u - l_u|$.*

Theorem 1 *Let us define $w = \alpha u$ where u is a random variable and α a scale factor. The supports of u and w are linked by:*

$$||\Omega_w|| = |\alpha|.||\Omega_u|| \ . \tag{5}$$

Proof: Let $\Omega_w = [l_w, r_w]$, then $u \in [\frac{l_w}{\alpha}, \frac{r_w}{\alpha}] = \Omega_u \doteq [l_u, r_u]$. Hence, $||\Omega_w|| = |\alpha|.||\Omega_u||$.

Theorem 2 *Assume that u and v are independent random variables. Then:*

$$||\Omega_{u+v}|| = ||\Omega_u|| + ||\Omega_v|| \ . \tag{6}$$

This result is an immediate consequence of the fact that the distribution p_{u+v} is the convolution product of p_u and p_v [4].

Corollary 1 *If $y \doteq \alpha u + \beta v$ where u and v are two independent random variables and α, β are real scale factors, the following expression holds for $||\Omega_y||$:*

$$||\Omega_y|| = |\alpha|.||\Omega_u|| + |\beta|.||\Omega_v|| \ . \tag{7}$$

The proof results from Theorem 1 and Theorem 2.

2.2. Finite sources

Definition 2 (Finite source) *Let u be a one-dimensional random variable with unbounded support: $\Omega_u =]-\infty, +\infty[$. The variable u is said to be* finite *if:*

$$\int_{l_u}^{r_u} p_u(\tau) d\tau \simeq 1 \textit{ with } -\infty < l_u < r_u < +\infty \ . \tag{8}$$

Unfortunately, this definition does not allow to give precise values of the pseudo-bounds l_u and r_u; the measure of the support width cannot be properly defined using this definition. Nevertheless, in the case of finite sources, other definitions may be proposed, as explained in Section 4.

3. CONVEX CONTRAST: OUTPUT SWM

A well-known problem of gradient-based multimodal source separation is the existence of spurious maxima for usual BSS contrasts. This is e.g. the case when using the opposite of the output marginal entropy $-H(y_i)$ [5, 6, 7].

In order to avoid the existence of spurious minima, it is interesting to derive contrasts that are convex on each quadrant $\mathcal{Q}_p$. In Section 3.1, the convexity of the contrast function $-||\Omega_{y_i}||$ for $\theta \in \mathcal{Q}_p(\theta)$ and bounded sources is proven in all quadrants ($p \in \{1, 2, 3, 4\}$). An extension of the support width concept to finite sources is given in Section 3.2, and applied on a simple example.

3.1. Bounded sources

The convexity on $\mathcal{Q}_p$ property, which will be proven for bounded sources, implies that the maximum value of $-||\Omega_{y_i}||$ is reached for $\theta \in \{k\pi/2\}$. Combining eq. 4 and Corollary 1, the contrast $\mathcal{C}^2$ can be rewritten as:

$$\mathcal{C}^2(\theta) \doteq -||\Omega_{y_1}|| = -|\cos\theta|.||\Omega_{s_1}|| - |\sin\theta|.||\Omega_{s_2}|| \ . \tag{9}$$

where the superscript of $\mathcal{C}$ denotes the number of sources. Note that one could think that $\mathcal{C}^2$ is equivalent to $\overline{\mathcal{C}}^2 \doteq -|\cos\varphi|.||\Omega_{z_1}|| - |\sin\varphi|.||\Omega_{z_2}||$. In this case, one could compute by an algebraic way the angle φ that maximizes it. Unfortunately, since z_1 and z_2 are only uncorrelated and not statistically independent, they do not respect the necessary condition of Corollary 1 and $\overline{\mathcal{C}}^2$ cannot be used as a contrast for source separation (see Fig. 2). Hence, to evaluate $\mathcal{C}^2$, we have to compute $||\Omega_{y_1}||$ by estimating the bounds of y_1 and applying Definition 1.

Definition 3 *A function $\mathcal{C}(\xi)$ is said to be convex on $\mathcal{Q}_p$ if*

$$\mathcal{C}\Big(\lambda\xi_1 + (1-\lambda)\xi_2\Big) \leq \lambda\mathcal{C}(\xi_1) + (1-\lambda)\mathcal{C}(\xi_2) \tag{10}$$

for all $\xi_1, \xi_2 \in \mathcal{Q}_p$, for $0 \leq \lambda \leq 1$.

Theorem 3 (Convexity of $\mathcal{C}^2(\theta)$) *The function $\mathcal{C}^2(\theta)$ is convex on $\theta \in \mathcal{Q}_p$.*

Proof: A sufficient condition for ensuring that $\mathcal{C}^2(\theta)$ is convex on $\theta \in \mathcal{Q}_p$ is to prove that $\frac{\partial^2 \mathcal{C}^2(\theta)}{\partial \theta^2} > 0$ [8]. We focus on the $p = 1$ case. It is obvious that this condition is respected, since $\frac{\partial^2 \mathcal{C}^2(\theta)}{\partial \theta^2} = \cos\theta.||\Omega_{s_1}|| + \sin\theta.||\Omega_{s_2}|| > 0$. It can be shown easily that this result holds for other values of p.

As a consequence, the function $\mathcal{C}^2(\theta)$ is a convex contrast for BSS if the sources are zero-mean, white and independent.

Corollary 2 *If $\varphi^\star$ is the unmixing angle maximizing locally $\mathcal{C}^2(\theta)$, then $\theta^\star \doteq \phi + \varphi^\star = k\pi/2$.*

This result is an immediate consequence of the convexity of $\mathcal{C}^2(\theta)$ in a given quadrant $\mathcal{Q}_p$. By Corollary 2, if we derive a measure of $\mathcal{C}^2(\theta)$, it is possible to achieve BSS simply by maximizing $\mathcal{C}^2(\theta)$.

Note that the function $\mathcal{C}^2(\theta)$ has a minimum for $\theta = \arctan \frac{||\Omega_{s_2}||}{||\Omega_{s_1}||}$; its second derivative is always positive for $\theta \in \mathcal{Q}_p(\theta)\backslash\{k\pi/2\}$, and does not exist for $\theta = k\pi/2$.

3.2. SWM for finite sources

Section 3.1 proves that, in the case of bounded sources, it is possible to derive a convex contrast. The illustration of the extension of this result to finite sources is given here. The following example shows that maximizing an approximation $||\hat{\Omega}_{y_1}||$ of $||\Omega_{y_1}||$ allows to recover blindly the unknown mixing angle ϕ. According to numerical simulations, it exists a definition of bounds that allows one to preserve the convexity property of the approximated contrasts.

Let $\Upsilon(x) = \epsilon > 0$ be a constant function on Ω_{y_1} (with ϵ small enough). Assume that $x^r > 0$ and $x^l < 0$ correspond to two roots of the $\Delta(x) \doteq \Upsilon(x) - p_{y_1}(x)$ and for $x \geq x^r$ (resp. $x \leq x^l$) the function $\Delta(x)$ has no more roots (i.e. $\Delta(x) \neq 0$). Let us define x^r and x^l as the upper and lower pseudo-bounds of the support of y_1, respectively. Hence, according to this definition of pseudo-bounds, we can derive an under-estimation of the SWM of a finite signal:

$$\hat{\mathcal{C}}^2(\theta) = -||\hat{\Omega}_{y_1}|| \doteq -(x_r - x_l) \ . \qquad (11)$$

The contrasts $\mathcal{C}^2$, $\hat{\mathcal{C}}^2$, $-H(y_1)$ and the function $\overline{\mathcal{C}}^2$ are plotted versus the unmixing angle φ in Fig. 2 for the mixtures of Fig. 1(b) (the distributions p_{y_i} are estimated by the Parzen estimator with Gaussian kernels). The mixing angle has been taken equal to $\pi/3$ in this example.

4. EXTENSION TO MORE SOURCES

It has been shown in [3] that $\mathcal{C}^\Sigma \doteq -\sum_{i=1}^n \log||\Omega_{y_i}||$ is a global contrast to separate instantaneous and linear mixtures of n sources. In this section, we will show that $\mathcal{C}^n \doteq$

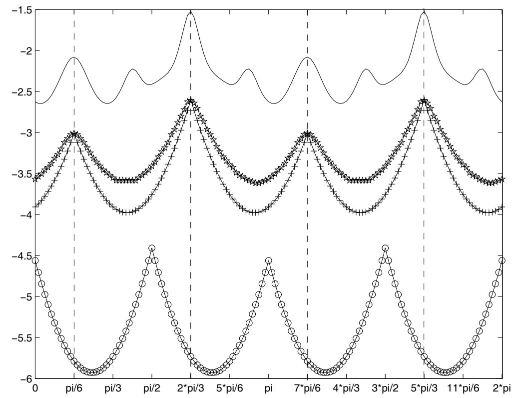


Fig. 2. Contrasts $\mathcal{C}^2$ (+), $\hat{\mathcal{C}}^2$ ($\star$), $-H(y_1)$ (-) and function $\overline{\mathcal{C}}^2$ ($\circ$) vs φ. The only relevant maxima are located at $\varphi = \pi/6 + k\pi/2 | k \in \mathbb{Z}$ ($\phi = \pi/3$);

$-||\Omega_{y_i}||$ is a convex contrast (without mixing maxima). For $n > 2$, using the associativity property of the convolution product, the SWM-based contrast $\mathcal{C}^n$ can be rewritten as:

$$\mathcal{C}^n(\mathbf{w}) \quad = \quad -\sum_{i=1}^{n} |w_i|.||\Omega_{s_i}||, \ \text{ s.t. } \sum_i w_i^2 = 1 \ ,(12)$$

where the w_i are the elements of a row of $\mathbf{W}$.

The two next propositions prove that i) the global maximum of $-||\Omega_{y_i}||$ is reached when only the j^{th} entry of w_i is non-zero where $j = \mathrm{argmin}_j \, ||\Omega_{s_j}||$ and ii) each maximum of $\mathcal{C}^n$ is relevant for source separation.

Proposition 1 $\arg\max_{\mathbf{w}} \mathcal{C}^n(\mathbf{w}) = \mathbf{e}_i$*, where* $(\mathbf{e}_i)_k = \delta_{k,i}$*,* $1 \leq k \leq n$*,* δ *is the Kronecker symbol, and* i *is such that* $||\Omega_{s_i}|| = \min_j \left\{ ||\Omega_{s_j}|| \right\}$.

Proof: Consider the weight vector $\mathbf{w}^p = [w_1^p, w_2^p, \ldots, w_n^p]$ where at least two elements are non-zero; without loss of generality, $w_1, w_2 > 0$, and $w_j \geq 0$ $(3 \leq j \leq n)$. Let us note for simplicity $a_i \doteq ||\Omega_{s_i}|| > 0$ and suppose that $a_1 \leq a_2$. Hence, it exists another weight vector $\mathbf{w}^q$ such that $\mathcal{C}^n(\mathbf{w}^q) > \mathcal{C}^n(\mathbf{w}^p)$.

Let $\mathbf{w}^q \doteq [\sqrt{(w_1^p)^2 + (w_2^p)^2}, 0, w_3^p, \ldots, w_n^p]$, then:

$$\begin{aligned}
(w_2^p)^2 a_1^2 &\leq (w_2^p)^2 a_2^2 \\
(w_1^p)^2 a_1^2 + (w_2^p)^2 a_1^2 &< (w_1^p)^2 a_1^2 + (w_2^p)^2 a_2^2 + \underbrace{2 w_1^p w_2^p a_1 a_2}_{>0} \\
a_1 \sqrt{(w_1^p)^2 + (w_2^p)^2} &< w_1^p a_1 + w_2^p a_2 \\
w_1^q a_1 + w_2^q a_2 &< w_1^p a_1 + w_2^p a_2 \qquad (13)
\end{aligned}$$

Hence, as $w_j^p = w_j^q$ for $j \geq 3$, if $w_1^q a_1 + w_2^q a_2 < w_1^p a_1 + w_2^p a_2$ then $\mathcal{C}^n(\mathbf{w}^q) > \mathcal{C}^n(\mathbf{w}^p)$. Iterating this result, we have that $\mathcal{C}^n(\mathbf{e}_i) \geq \mathcal{C}^n(\mathbf{w})$, for all $\mathbf{w}$, with equality if and only if $\mathbf{w} = \mathbf{e}_i$ and where i is such that $a_i = \min\{a_j\}$.

Proposition 2 *If* $\mathbf{w} \neq \mathbf{e}_k$*, then* $\mathcal{C}^n(\mathbf{w})$ *cannot be a local maximum.*

Proof: Because of the lack of space, we will give only the sketch of the proof. Consider the infinitesimal vector $\Delta^A \mathbf{w}$. If $\mathcal{C}^n(\mathbf{w} + \Delta^A \mathbf{w}) > \mathcal{C}^n(\mathbf{w})$, the proposition is proven. Algebraic manipulations show that when $\mathcal{C}^n(\mathbf{w} + \Delta^A \mathbf{w}) \leq \mathcal{C}^n(\mathbf{w})$, it exists another infinitesimal vector $\Delta^B \mathbf{w}$ such that in this case, we have $\mathcal{C}^n(\mathbf{w} + \Delta^B \mathbf{w}) > \mathcal{C}^n(\mathbf{w})$. One can prove this result using $\Delta^A \mathbf{w}_i = \delta w$ and $\Delta^A \mathbf{w}_j$ such that $\sum_i (\mathbf{w} + \Delta^A \mathbf{w})_i^2 = 1$ with $\Delta^A \mathbf{w}_{r \neq i,j} = 0$; δw denotes an infinitesimal number. Similarly, we can choose $\Delta^B \mathbf{w}_i = -\delta w$ and $\sum_i (\mathbf{w} + \Delta^B \mathbf{w})_i^2 = 1$ with $\Delta^B \mathbf{w}_{r \neq i,j} = 0$. In summary, if $\mathbf{w} \neq \mathbf{e}_k$, it always exists a small vector $\Delta \mathbf{w}$ for which $\mathcal{C}^n(\mathbf{w} + \Delta \mathbf{w}) > \mathcal{C}^n(\mathbf{w})$ respecting $\sum_i (\mathbf{w} + \Delta \mathbf{w})_i^2 = 1$. Note that there is no guarantee that all local maxima are global ones: local maxima may exist (see e.g. Fig. 2). Nevertheless, this does not matter, since i) we know by Proposition 2 that a weight vector $\mathbf{w} \neq \mathbf{e}_k$ cannot be a maximum of $\mathcal{C}^n$ and ii) each $\mathbf{w} = \mathbf{e}_k$ $(1 \leq k \leq n)$ corresponds to the extraction of one source!

Fig. 3 shows projections of the manifold of $\hat{\mathcal{C}}^3 \doteq -||\hat{\Omega}_{y_1}||$ on the w_1, w_2 plane for two triplets of source support widths. The global maxima (darkest areas) correspond to the weight vector $\mathbf{w} = \mathbf{e}_i$ where $i = \mathrm{argmin}_j \, ||\Omega_{s_i}||$. Only the weight corresponding to the source with minimum SWM is non zero. This weight is equal to one because of the $\sum_{i=1}^3 w_i^2 = 1$ constraint. All the maxima correspond to $\mathbf{w} = \mathbf{e}_k$.

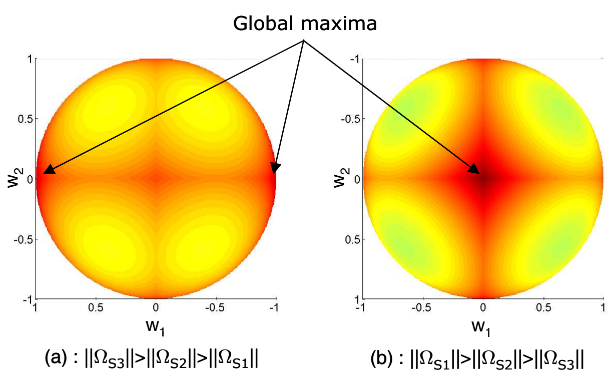


Fig. 3. Manifold of the contrast function $\hat{\mathcal{C}}^3$ projected on the 2D space w_1, w_2.

5. DISCUSSION AND CONCLUSION

This paper presents a contrast function to separate bounded sources, based on a support width measure (SWM). The contrast is shown to be *convex*, which means that non-mixing maxima can be reached, in theory, by gradient-ascent based algorithms.

For finite sources (with virtually unbounded support), approximations can be used to define pseudo-bounds of the support, using a constant function $\Upsilon(x)$. Simulations show that the convexity property is preserved with these approximations.

For $n = 2$, ICA algorithms can be derived to maximize the contrast function or its approximation by adding or subtracting small variations $\Delta\varphi$ to the unmixing angle φ.

In future work, the SWM method will be related to other algorithms exploiting the scatter plot boundaries, like geometric ICA. The bound estimation of a distribution may be difficult when few samples are available in these area of the distribution, and should be deeper investigated. In particular, it should be interesting to study the influence of $\Upsilon(x)$ on the convexity of $\hat{\mathcal{C}}^n$, and mixtures with additive noise. Other functions $\Upsilon(x)$ can be seen as a non-Gaussianity measure, and seem to be quite efficient in some cases (see e.g. [5], where $\Upsilon(x)$ is the normalized Gaussian function and is used for the extraction of multimodal sources). For the $n > 2$ case, it has been shown that $\mathcal{C}^n$ is also a convex contrast. However, in this case the BSS problem is not equivalent to recovering a single unmixing angle, and other ICA algorithms have to be derived.

6. REFERENCES

[1] A. Cichoki and S.-I. Amari, *Adaptive blind signal and image processing*, John Wiley & Sons, England, 2002.

[2] J.-F. Cardoso and P. Comon, "Independent component analysis, a survey of some algebraic methods," *International Symposium Circuits and Systems*, vol. 2, pp. 93–96, 1996.

[3] D.-T. Pham, "Blind separation of instantenaous mixtrures of sources based on order statistics," *IEEE Trans. on Signal Processing*, vol. 48, no. 2, pp. 363–375, 2000.

[4] I.I. Hirschman and D.V. Widder, *The convolution transform*, Princeton University Press, 1955.

[5] F. Vrins and M. Verleysen, "Information theoretic vs cumulant-based contrasts for multimodal source separation," *To appear in IEEE Sig. Proc. Letters.*

[6] E. G. Learned-Miller and J. W. Fisher III, "ICA using spacings estimates of entropy," *Journal of Machine Learning Research*, vol. 4, pp. 1271–1295, 2003.

[7] F. Vrins and M. Verleysen, "On the entropy minimization of a linear mixture of variables for source separation," *To appear in Signal Processing*.

[8] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley and sons, 1991.